

Basic Definitions

- A **graph** is a way of specifying relationships among a collection of items.
 - A **graph** consists of a set of objects called **nodes** (**vertices**)
 - With certain pairs of these objects connected by links called **edges** (**links**)
 - Two nodes are **neighbors** if they are connected by an **edge**

Directed vs. Undirected

- Edges have no orientation (e.g, Facebook)
- Edges have orientation (e.g., Twitter)

Weighted Graphs

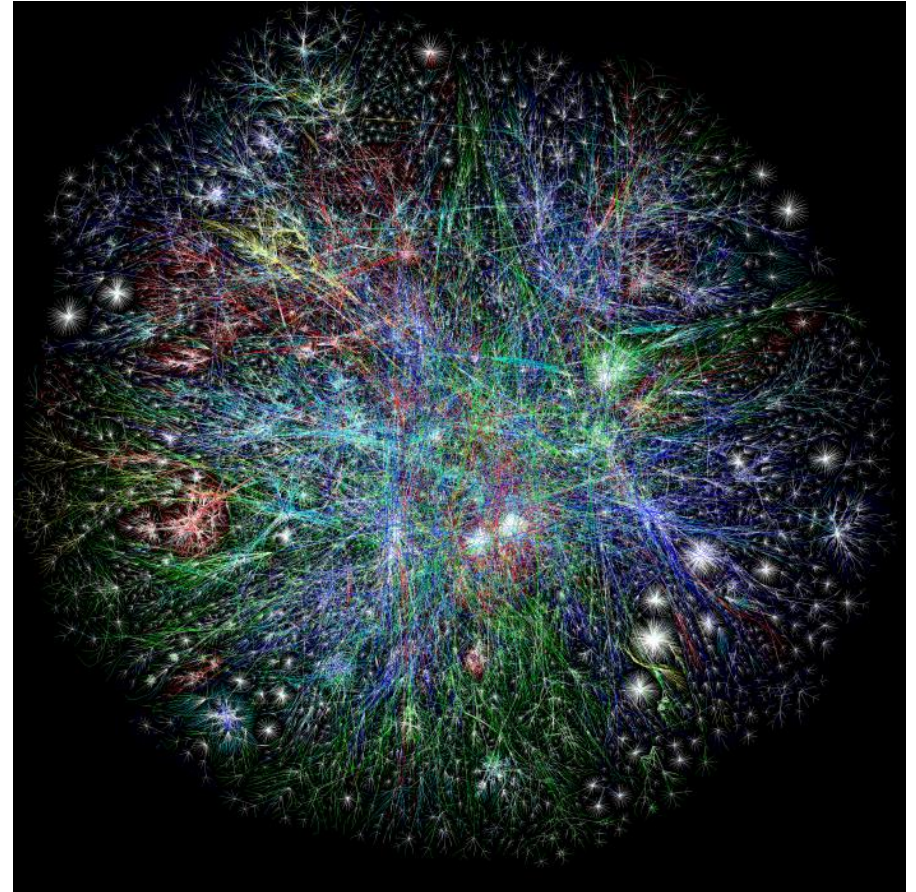
- In a *weighted* graph every edge has an associated *weight* with it
 - What could weights represent?
 - E.g, distance, cost, frequency etc

Bipartite graph

- **Bipartite graph** (or **bigraph**) is a graph whose vertices can be divided into two disjoint sets U and V and such that every edge connects a vertex in U to one in V .
 - Complete Bipartite Graph

Main Concepts

- Paths
- Cycles
- Connectivity
- (Giant) Components
- Distance



Paths

- A **path** in a graph is a sequence of nodes with the property that each consecutive pair in the sequence is connected by an edge
 - In a **simple path** nodes do not repeat
 - Sometimes can be referred to **Walks** and **Paths**

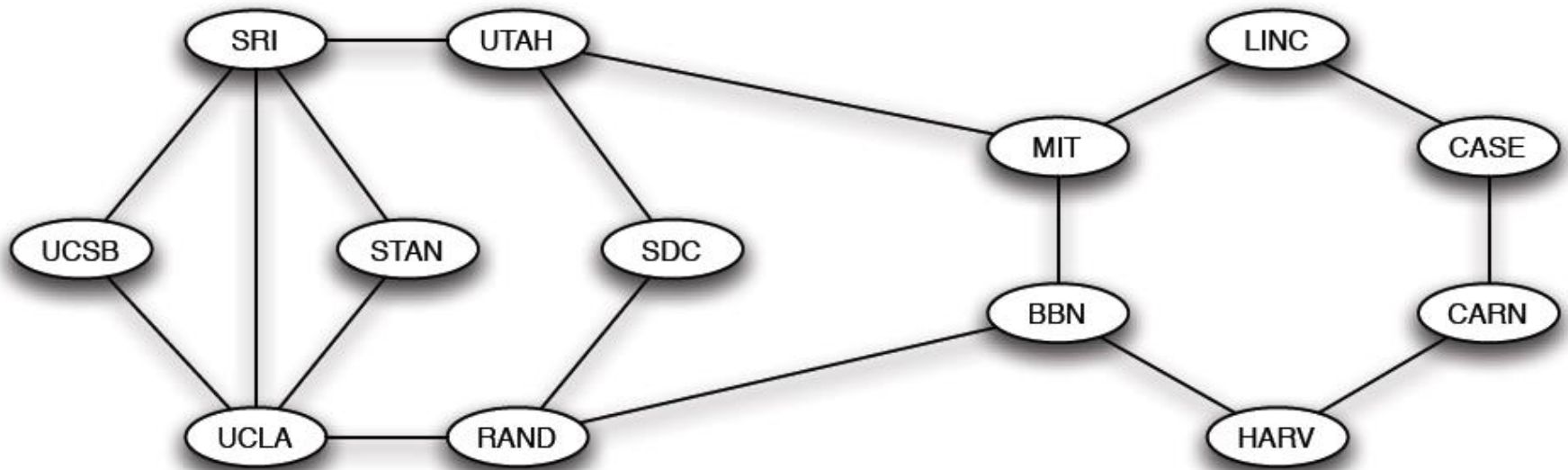


Figure 2.3: An alternate drawing of the 13-node Internet graph from December 1970.

Diameter

- Graph **diameter** can be measured as the
 - Largest shortest path
 - Any issue with that?
 - Average of shortest paths among all the nodes
 - Less sensitive to outliers
- Examples (what are the diameters of these graphs?)
 - $O(N)$ $O(\log N)$
 - WWW: 3,1, FB: 4,74, Co-authorship graphs: 5-10
 - Usually diameter is considered “small” if it is $O(\log N)$.
 - We’ll see later that it will depend on the degree of the network

Cycles

- A **cycle** is a closed path with at least three edges
 - All nodes are distinct except the first and the last
 - Why cycles are useful?
 - Every edge in 1970 network belongs to a cycle: design choice for making network connected even if one link failed.

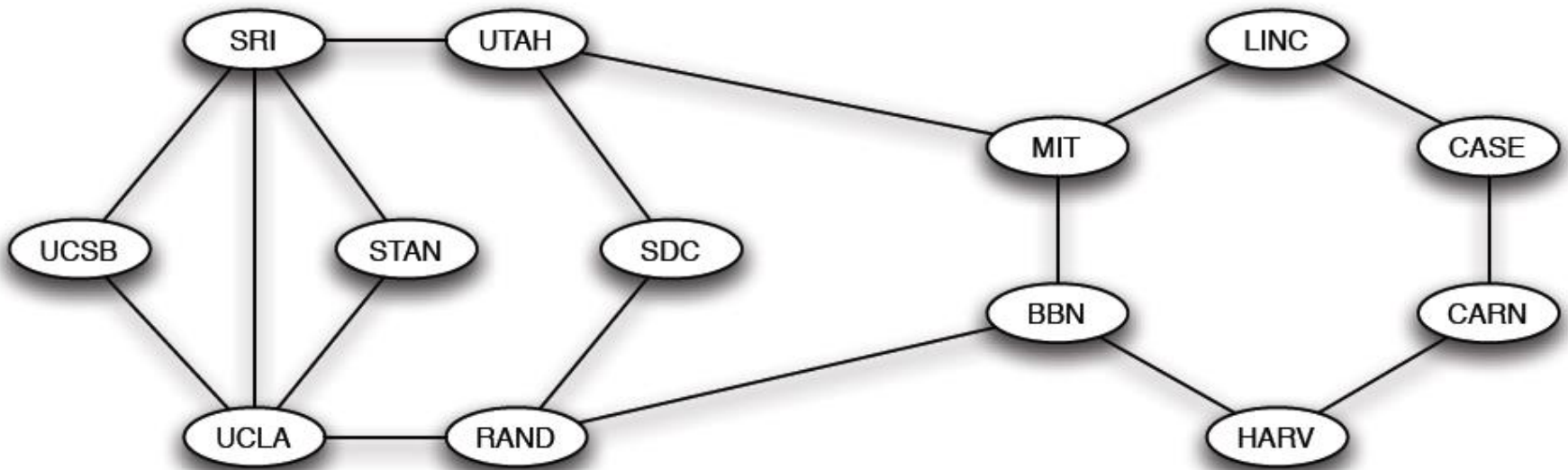


Figure 2.3: An alternate drawing of the 13-node Internet graph from December 1970.

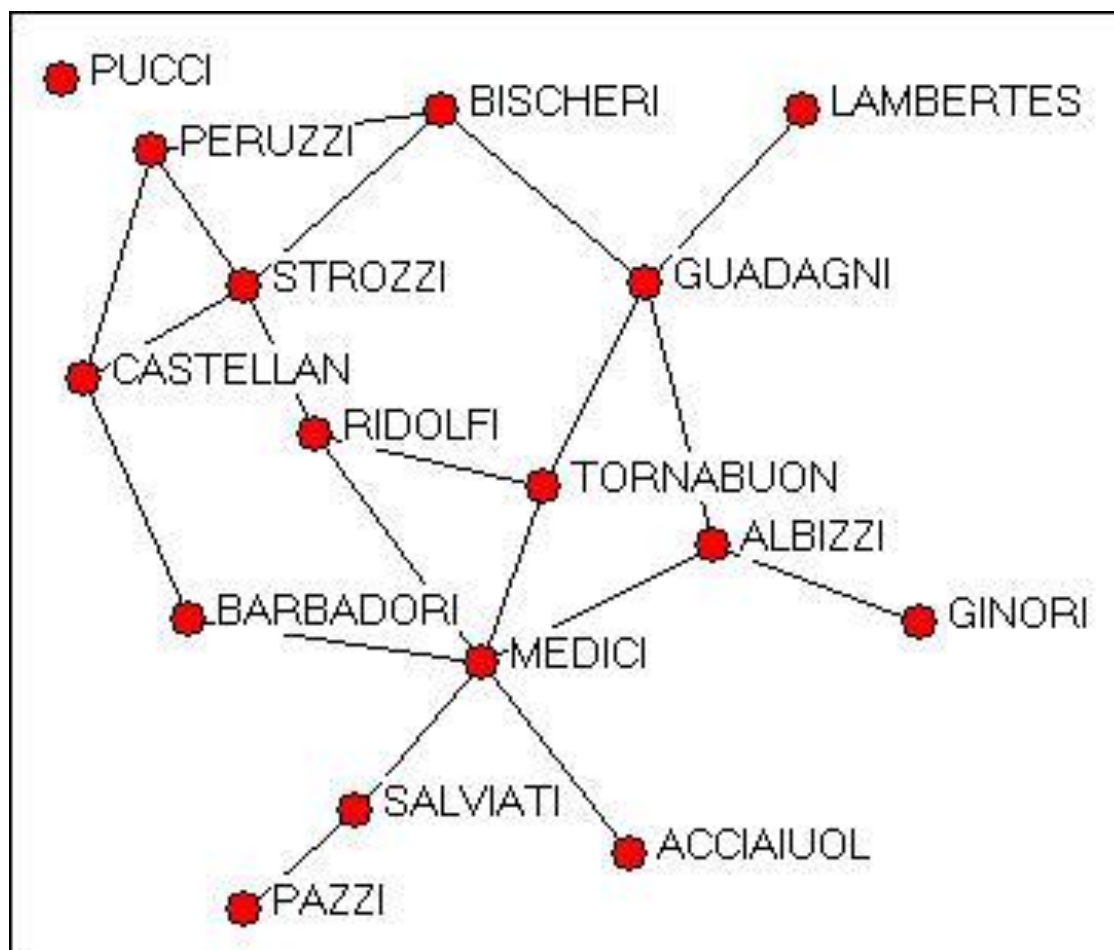
Connectivity

- A graph is ***connected*** if for every pair of nodes there is a path between them.
- A ***disconnected*** graph is made of at least two connected sub-graphs (components)

Connectivity (cont.)

- Local bridge
 - AB edge is a local bridge if A and B have no neighbors in common, but there exist another path from A to B.
- Embeddedness of the edge
 - number of mutual friends that the endpoints of the edge have in common.

Padgett' s Florentine Families



Betweenness Centrality

- Intuition: *how many pairs of nodes have a shortest path through you?*
- Betweenness centrality:
 - $C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$
 - σ_{st} number of shortest paths between s and t
 - $\sigma_{st}(v)$ number of shortest paths between s and t via v.
- Can be normalized:
 - $C'_B(v) = \frac{C_B(v)}{(n-1)(n-2)/2}$

Examples

Connectivity (directed graphs)

- A directed graph is ***strongly connected*** if for every pair of nodes there is a path between them. A ***weakly connected*** graph is connected if we disregard edge directions

Giant Component

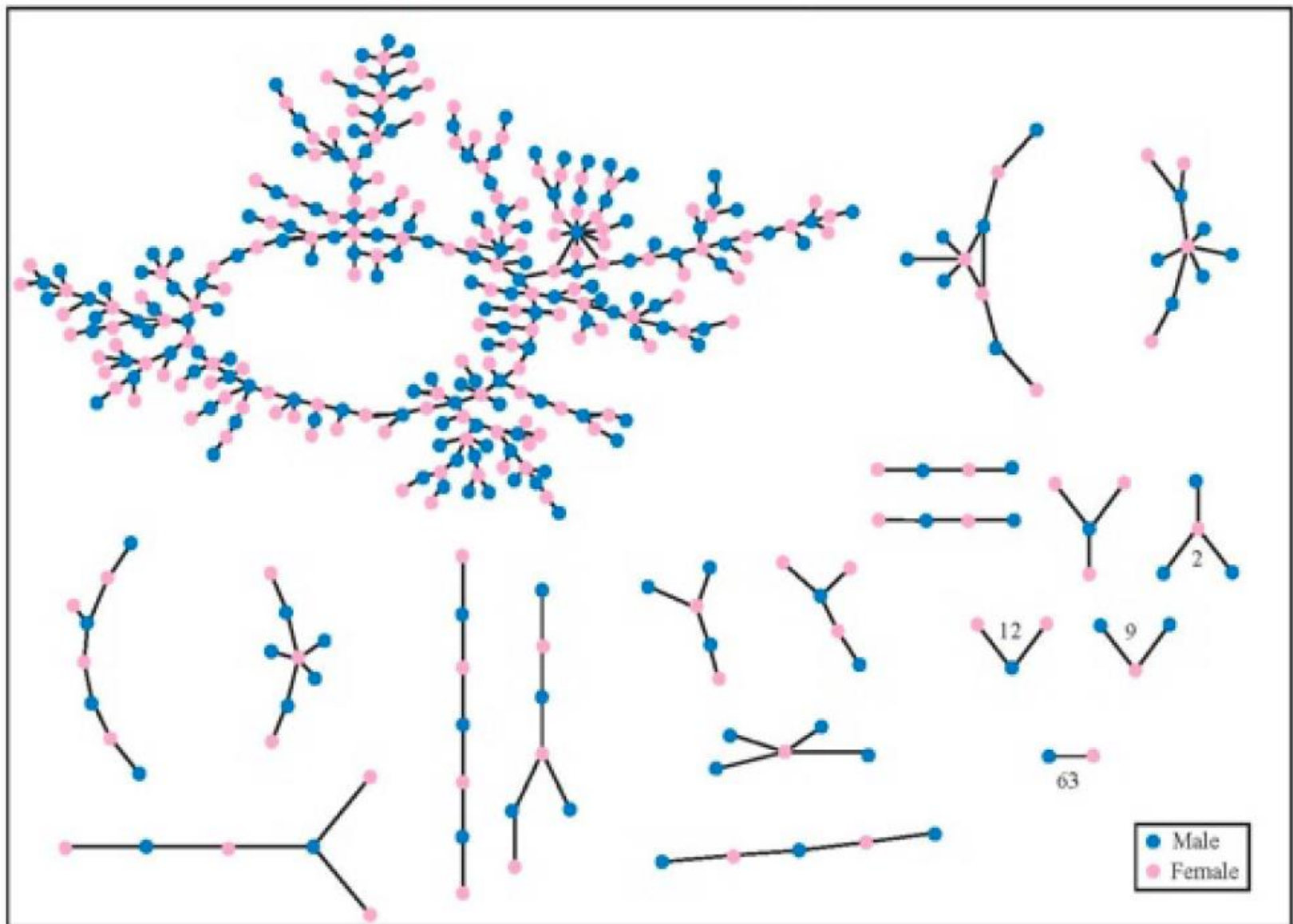
- A ***connected component*** of a graph is a subset of the nodes such that:
 - (i) every node in the subset has a path to every other;
 - and (ii) the subset is not part of some larger set with the property that every node can reach every other.
- ***Giant component***: a connected component with the largest number of nodes

Giant Components

- Real World networks often contain only a specific number of **largest** components that are **similar** in size.
 - Think what could this number be?
 - A: 1
 - B: 2
 - C: 3
 - D: 4 to 10
 - E: 11 to 99
 - F: 100 and more
- Real world networks often contain only **one giant component**

why?

Intuition on why there can't be 2 giant
components



- <http://ccl.northwestern.edu/netlogo/models/GiantComponent>

Clustering coefficient

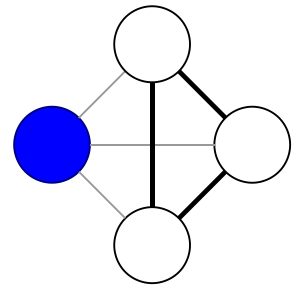
- **Local clustering coefficient** $C(v)$ of vertex v is given by the fraction of:

$$C(v) = \frac{e(v)}{\deg(v)(\deg(v) - 1) / 2}$$

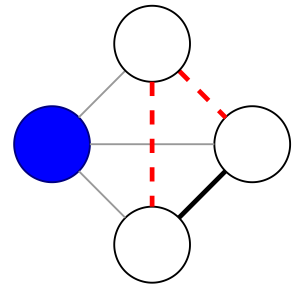
where $e(v)$ denotes the links between the vertices within the neighborhood of v

- **Network average clustering coefficient** $\tilde{C}$ is given by the fraction of:

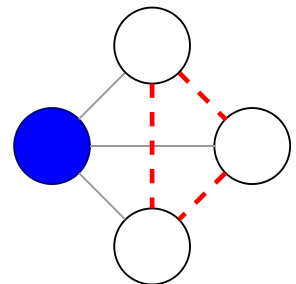
$$\tilde{C} = \frac{1}{N} \sum_{i=1}^N C(i)$$



$$c = 1$$



$$c = 1/3$$



$$c = 0$$

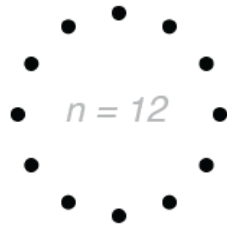
How to interpret clustering coef.?

- Clustering coefficient denotes what is the fraction of your neighbors are neighbors themselves
- Compare to a purely random chance that the “triangles” form.
- Edge density of a network: $p = \frac{E}{0,5 * N(N - 1)}$
 - E is total number of edges
 - P is the probability that two nodes are connected in a random graph
- If $C(G) \gg p$ then we can claim that the **graph is clustered**

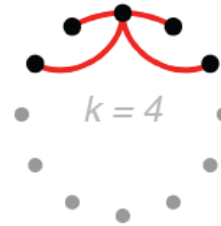
Examples

- Regular graph with degree k connected to nearest neighbors

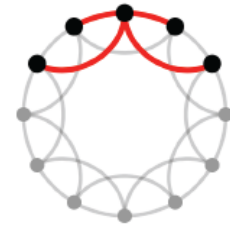
We start with a ring of n vertices



where each vertex is connected to its k nearest neighbors



like so.



- What's clustering coefficient when $k=4$?
 - Possible neighbor friendships: 6
 - Actual friendships: 3
- Clustering coef $3/6=0,5$
- Compare it with random graph?

- Assume we have a graph with $N=1\text{bn}$, and avg degree of 100. We measure avg clustering coefficient and find it to be $C(G)=0.0001$.
- Can we call this graph clustered?